한국통계학회논문집 제2권 2호
1995년 12월 pp. 376-386

Minimum Mean Squared Error Invariant Designs for Polynomial Approximation

Joong-Yang Park[1]

Abstract

Designs for polynomial approximation to the unknown response function are considered. Optimality criteria are monotone functions of the mean squared error matrix of the least squares estimator. They correspond to the classical A-, D-, G- and Q-optimalities. Optimal first order designs are chosen from the invariant designs and then compared with optimal second order designs.

1. Introduction

A response function $\eta(\boldsymbol{x})$ is the relationship between the expectation of the response variable y and k independent variables $\boldsymbol{x}=(x_1, x_2, \cdots, x_k)$. Suppose that we wish to design an experiment for investigating the response function over the experimental region $S=\{\boldsymbol{x} \mid \boldsymbol{x}'\boldsymbol{x} \leq 1\}$. In practice the response function is either very complicated or unknown. An experiment is usually designed for a multiple linear regression model for $\eta(\boldsymbol{x})$, which can be written as

$$y=f(\boldsymbol{x})'\boldsymbol{\beta}+\varepsilon, \tag{1}$$

where $f(\boldsymbol{x})$ is a vector of functions of $\boldsymbol{x}$, $\boldsymbol{\beta}$ is an unknown parameter vector which is to be estimated and ε is an error term with mean 0 and variance σ^2. This approach always causes some concerns about bias due to departures from model (1). An alternative design strategy is to design an experiment for model (1) so that the precision of the least squares estimator of $\boldsymbol{\beta}$ is robust against the departures. We thus consider the optimality criteria which are monotone functions of the mean squared error (MSE) matrix of the least squares estimator. Since the MSE matrix depends on the model departures, most of the previous works on robust designs assume that the true model is

$$y=f(\boldsymbol{x})'\boldsymbol{\beta}+z(\boldsymbol{x})+\varepsilon, \tag{2}$$

1) Department of Statistics, Gyeongsang National University, Chinju, 660-701, KOREA.

where $z(\boldsymbol{x})$ belongs to some specified class $\mathcal{F}$ of functions of $\boldsymbol{x}$. Markus and Sacks (1976), Sacks and Ylvisaker (1978), Pesotchinsky (1982), Li and Notz (1982) and Li (1984) take

$$\mathcal{F}=\{z(\boldsymbol{x}) \mid |z(\boldsymbol{x})| \leq \psi(\boldsymbol{x}) \text{ for all } \boldsymbol{x} \in S\}$$

with various assumptions about ψ. Huber (1975) and Weins (1992) takes

$$\mathcal{F}=\left\{z(\boldsymbol{x}) \mid \int_S z(\boldsymbol{x})^2 d\boldsymbol{x} \leq \zeta \text{ and } \int_S f(\boldsymbol{x}) z(\boldsymbol{x}) d\boldsymbol{x}=\mathbf{0}\right\},$$

where ζ is assumed known. The unknown response function is mostly approximated by some low order polynomial model. Assuming that $f(\boldsymbol{x})'\boldsymbol{\beta}$is a polynomial of order d, $z(\boldsymbol{x})$ can be regarded as the remainder consisting of multiple monomials of order $(d+1)$. That is, $f(\boldsymbol{x})$ is the vector of p multiple monomials $\prod_{i=1}^{k} x_i^{\alpha_i}$ up to degree d where $\sum_{i=1}^{k}\alpha_i \leq d$ and $p=\binom{k+d}{d}$,

$z(\boldsymbol{x})=h(\boldsymbol{x})'\boldsymbol{\gamma}$ where $h(\boldsymbol{x})$ is the vector of multiple monomials of degree $(d+1)$ and $\boldsymbol{\gamma}$ is the vector of parameters corresponding to $h(\boldsymbol{x})$. If $\eta(\boldsymbol{x})$ and its first d derivatives are continuous over S and $(d+1)$th derivative of $\eta(\boldsymbol{x})$ exists, $|z(\boldsymbol{x})|$ is bounded above. This paper therefore takes

$$\mathcal{F}=\{h(\boldsymbol{x})'\boldsymbol{\gamma} \mid |h(\boldsymbol{x})'\boldsymbol{\gamma}| \leq \delta \text{ for all } \boldsymbol{x} \in S\}, \tag{3}$$

where δ is assumed known.

We consider the problem of constructing the optimal designs for polynomial regression model $y=f(\boldsymbol{x})'\boldsymbol{\beta}+\varepsilon$ under the assumption that the true model is $y=f(\boldsymbol{x})'\boldsymbol{\beta}+h(\boldsymbol{x})'\boldsymbol{\gamma}+\varepsilon$ where $h(\boldsymbol{x})'\boldsymbol{\gamma}$ belongs to class (3). A similar design problem was researched by Box and Draper (1959), who advocated the average MSE (AMSE) criterion for the estimation of mean response $\eta(\boldsymbol{x})$ and were followed by Box and Draper (1963) and Draper and Lawrence (1965). These studies require an estimate of the magnitude of $\boldsymbol{\gamma}$. In practice it is more plausible to estimate the upper bound of the model departure, δ in (3), than to estimate $\boldsymbol{\gamma}$. In order to overcome the dependency of AMSE on $\boldsymbol{\gamma}$, minimum bias designs and designs for minimum bias estimation have been studied by Karson, Manson and Hader (1969), Myers and Lahoda (1975), Draper and Sanders (1988) and Park (1990). However, such designs place emphasis on the bias component of AMSE. As mentioned above, our design objective is the estimation of parameters. MSE versions of the classical A-, D-, G- and Q-optimalities are employed as the optimality criteria. Section 2 presents the definitions of the optimality criteria and some results on the optimal design problem. Optimal first order designs are derived in Section 3 and their performances are investigated in Section 4.

2. Optimal design problem

The design problem is to choose n, not necessarily distinct, design points in S. Then a design ξ can be regarded as a probability measure on S defined by $N(\boldsymbol{x})/n$ where $N(\boldsymbol{x})$ is the repetition of the design point $\boldsymbol{x}$. We extend the definition of an experimental design to include all probability measures on S. This is the so-called approximate design theory. Henceforth ξ and Ξ will denote an arbitrary probability measure and the set of all probability measures on S. The MSE matrix of the least squares estimator of $\boldsymbol{\beta}$, standardized with respect to n and σ^2, for given design ξ is

$$M(\boldsymbol{\gamma},\xi)=M_1^{-1}(\xi)+\frac{n}{\sigma^2}M_1^{-1}(\xi)M_2(\xi)\boldsymbol{\gamma}\boldsymbol{\gamma}'M_2'(\xi)M_1^{-1}(\xi),$$

where $M_1(\xi)=\int_S f(\boldsymbol{x})f(\boldsymbol{x})'\,d\xi(\boldsymbol{x})$ and $M_2(\xi)=\int_S f(\boldsymbol{x})h(\boldsymbol{x})'\,d\xi(\boldsymbol{x})$. In selecting a design, it is desirable to minimize $M(\boldsymbol{\gamma},\xi)$ in the sense of a reasonable optimality criterion φ. We assume that φ is a convex and increasing function defined on the set $\mathcal{B}$ of all MSE matrices $M(\boldsymbol{\gamma},\xi)$ when $h(\boldsymbol{x})'\boldsymbol{\gamma}$ and ξ range over $\mathcal{F}$ and Ξ. Some of such criteria under our considerations are $tr(M(\boldsymbol{\gamma},\xi))$, $det(M(\boldsymbol{\gamma},\xi))$, $\sup_S d(\boldsymbol{x}|\boldsymbol{\gamma},\xi)$ and $\int_S d(\boldsymbol{x}|\boldsymbol{\gamma},\xi)d\boldsymbol{x}$, where tr and det are the trace and determinant and $d(\boldsymbol{x}|\boldsymbol{\gamma},\xi)=f(\boldsymbol{x})'M(\boldsymbol{\gamma},\xi)f(\boldsymbol{x})$. They are respectively denoted by φ_A, φ_D, φ_G and φ_Q and simply referred to as A-, D-, G- and Q-optimalities in this paper. We suggest a minimax design ξ^*, i.e., one for which $\sup_{\mathcal{F}}\varphi(M(\boldsymbol{\gamma},\xi^*))=\inf_{\Xi}\sup_{\mathcal{F}}\varphi(M(\boldsymbol{\gamma},\xi))$. Such designs will be called the minimax MSE (MMSE) designs.

It is a heavy task to find the MMSE designs. We need to reduce the optimization problem to a manageable problem. First note that

$$\max_S\{h(\boldsymbol{x})'\boldsymbol{\gamma}\}^2=h_{\max}\,ch_{\max}(\boldsymbol{\gamma}\boldsymbol{\gamma}')=h_{\max}\boldsymbol{\gamma}'\boldsymbol{\gamma},$$

where $h_{\max}=\max_S h(\boldsymbol{x})'h(\boldsymbol{x})$ and $ch_{\max}$ denotes the maximum characteristic root. Since

$h(\mathbf{x})$ is specified and $h_{\max}$ can be evaluated, class (3) is equivalent to $\Gamma=\{\boldsymbol{\gamma} \mid \boldsymbol{\gamma}'\boldsymbol{\gamma} \le \tilde{\delta}\}$ where $\tilde{\delta}=\delta^2/h_{\max}$. Next consider $\boldsymbol{\gamma}\in\Gamma$ such that $\boldsymbol{\gamma}'\boldsymbol{\gamma}=w^2\tilde{\delta}<\tilde{\delta}$. Then $w^{-1}\boldsymbol{\gamma}\in\Gamma$, $w^{-2}\boldsymbol{\gamma}'\boldsymbol{\gamma}=\tilde{\delta}$ and $M(w^{-1}\boldsymbol{\gamma},\xi)-M(\boldsymbol{\gamma},\xi)\ge 0$ in the sense of positive definiteness. The monotonicity of φ ensures that it suffices to restrict our attention to $\Gamma_0=\{\boldsymbol{\gamma} \mid \boldsymbol{\gamma}'\boldsymbol{\gamma}=\tilde{\delta}\}$. If $\mathscr{B}$ is convex with respect to ξ, the minimax designs can be found from the invariant designs. (Refer to Heiligers (1991, 1992) and Heiligers and Schneider (1992).) Unfortunately due to the lack of convexity of $\mathscr{B}$, it is almost impossible to solve the optimization problem directly. Noting that $\mathscr{F}$ and Γ_0 are invariant with respect to orthogonal transformations and following the discussion in Kiefer (1959), we suggest that the MMSE designs are selected from Ξ_0, the set of invariant designs. The optimization problem is now reduced to the minimization of $\sup_{\Gamma_0}\varphi(M(\gamma,\xi))$ with respect to $\xi\in\Xi_0$. In the following section we will derive the MMSE invariant designs for $d=1$.

3. MMSE invariant first order designs

This section supposes $d=1$. That is, $f(\mathbf{x})'\boldsymbol{\beta}$ and $f(\mathbf{x})'\boldsymbol{\beta}+h(\mathbf{x})'\boldsymbol{\gamma}$ are respectively first and second order polynomials for which

$$f(\mathbf{x}) = (1, x_1, \cdots, x_k)',$$

$$h(\mathbf{x}) = (x_1^2, \cdots, x_k^2, x_1x_2, \cdots, x_{k-1}x_k)',$$

$$\boldsymbol{\beta} = (1, \beta_1, \cdots, \beta_k)'$$

and

$$\boldsymbol{\gamma} = (\gamma_{11}, \cdots, \gamma_{kk}, \gamma_{12}, \cdots, \gamma_{(k-1)k})'. \qquad (4)$$

$M_i(\xi)$'s of invariant first order designs are given by

$$M_1(\xi)=\begin{pmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & \mu_2\mathbf{I}_k \end{pmatrix} \text{ and } M_2(\xi)=\begin{pmatrix} \mu_2\mathbf{1}_k' & \mathbf{0}' \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mu_2=\int_S x_i^2 d\xi$, $\mathbf{I}_k$ is the identity matrix of order k and $\mathbf{1}_k$ is the vector of ones. Then $M(\gamma,\xi)$ is obtained as

$$\begin{pmatrix} 1 & 0' \\ 0 & \mu_2 I_k \end{pmatrix} + n\sigma^{-2} \begin{pmatrix} \mu_2^2(\sum\gamma_{ii})^2 & 0' \\ 0 & 0 \end{pmatrix}.$$

The MSE matrix for invariant first order designs is characterized by μ_2. We thus write $\varphi(\boldsymbol{\gamma}, \mu_2)$ for $\varphi(M(\boldsymbol{\gamma}, \xi))$. It can be verified that

$$\varphi_A(\boldsymbol{\gamma}, \mu_2) = 1 + k\mu_2^{-1} + n\sigma^{-2}\mu_2^2(\sum\gamma_{ii})^2,$$

$$\varphi_D(\boldsymbol{\gamma}, \mu_2) = \mu_2^{-k} + n\sigma^{-2}\mu_2^{2-k}(\sum\gamma_{ii})^2,$$

$$\varphi_G(\boldsymbol{\gamma}, \mu_2) = 1 + \mu_2^{-1} + n\sigma^{-2}\mu_2^2(\sum\gamma_{ii})^2$$

and

$$\varphi_Q(\boldsymbol{\gamma}, \mu_2) = 1 + k(k+2)^{-1}\mu_2^{-1} + n\sigma^{-2}\mu_2^2(\sum\gamma_{ii})^2.$$

Therefore $\sup_{\Gamma_0}\varphi$ is obtained by maximizing $(\sum\gamma_{ii})^2$ subject to $\boldsymbol{\gamma}'\boldsymbol{\gamma} = \tilde{\delta}$. The maximum of $(\sum\gamma_{ii})^2$ is obtained as $k\tilde{\delta}$ by the method of Lagrange multiplier. Let $\nu = n\tilde{\delta}/\sigma^2$, which can be interpreted as the relative importance of bias versus variance. Assuming that ν is given, we have

$$\sup_{\Gamma_0}\varphi_A(\boldsymbol{\gamma}, \mu_2) = 1 + k\mu_2^{-1} + k\nu\mu_2^2,$$

$$\sup_{\Gamma_0}\varphi_D(\boldsymbol{\gamma}, \mu_2) = \mu_2^{-k} + k\nu\mu_2^{2-k},$$

$$\sup_{\Gamma_0}\varphi_G(\boldsymbol{\gamma}, \mu_2) = 1 + k\mu_2^{-1} + k\nu\mu_2^2$$

and

$$\sup_{\Gamma_0}\varphi_Q(\boldsymbol{\gamma}, \mu_2) = 1 + k(k+2)^{-1}\mu_2^{-1} + k\nu\mu_2^2$$

Convexity of $\sup_{\Gamma_0}\varphi(\boldsymbol{\gamma}, \mu_2)$ is apparent. Optimal values of μ_2 are therefore obtained by equating the derivative of each $\sup_{\Gamma_0}\varphi(\boldsymbol{\gamma}, \mu_2)$ with respect to μ_2 to zero and considering the moment condition $0 < \mu_2 \le k^{-1}$ as

$$\mu_{2A} = \min\{k^{-1}, (2\nu)^{-1/3}\},$$

$$\mu_{2D} = k^{-1},$$

$$\mu_{2G} = \min\{k^{-1}, (2k\nu)^{-1/3}\}$$

and

$$\mu_{2Q} = \min\{k^{-1}, (2(k+2)\nu)^{-1/3}\}. \tag{5}$$

It is interesting that D-optimal MMSE invariant designs do not depend on ν. Let ν_A^*, ν_G^* and ν_Q^* denote the maximums of ν such that the corresponding optimal value of μ_2 becomes k^{-1} and independent of ν. If ν is larger than the maximum, optimal value of μ_2 decreases as ν increases. Table 1 presents the maximums for $2 \leq k \leq 8$. It is evident that G- and Q -optimal MMSE designs are relatively more sensitive to ν than A- and D-optimal MMSE designs. The MMSE invariant first order designs can be constructed as 2^{k-t} (fractional) factorial designs consisting of the vertices and center point of S. Weights for each vertex and the center point are respectively $\mu_2 2^{t-k}$ and $(1-\mu_2)$.

Table 1. Values of ν_A^*, ν_G^* and ν_Q^*.

k	ν_A^*	ν_G^*	ν_Q^*
2	4.0000	2.0000	1.0000
3	13.5000	4.5000	2.7000
4	32.0000	8.0000	5.3333
5	62.5000	12.5000	8.9286
6	108.0000	18.5000	13.5000
7	171.5000	24.5000	19.0556
8	256.0000	32.0000	25.6000

4. Performance of MMSE invariant first order designs

In this section we study the performance of the MMSE invariant first order designs derived in Section 3. If the true model is a first order polynomial, then the optimal value of μ_2 is k^{-1} irrespective of the optimality criterion. This results from substituting ν in (5) with zero. Therefore, as long as the incorrectly assumed value of ν is not larger than the maximum of Table 1, the MMSE invariant first order designs coincide with the optimal designs for the true first order polynomial model. When a first order polynomial model is assumed and model departure caused by second order terms is concerned about, either the MMSE invariant first order designs or the optimal second order designs can be used. Thus it is necessary to compare the performances of these designs.

For investigating the performance of a design, Vining and Myers (1991) suggested plotting maximum and average MSE of the estimated response at every sphere in S. However, in order to examine the overall performances of the MMSE invariant first order designs, we consider maximum and average MSE of the estimated response over the experimental region S. Let $g(\boldsymbol{x}) = (f(\boldsymbol{x})', h(\boldsymbol{x})')'$ and $\widetilde{M}_1(\xi) = \int_S g(\boldsymbol{x})g(\boldsymbol{x})'\, d\xi(\boldsymbol{x})$, where $f(\boldsymbol{x})$ and $h(\boldsymbol{x})$ are given by (4). Denoting by $\xi^*_{A_2}$ the A-optimal second order designs minimizing $tr(\widetilde{M}_1(\xi))$, we define the overall efficiencies of the A-optimal MMSE invariant first order design $\xi^*_{A_1}$ relative to $\xi^*_{A_2}$ as

$$\text{eff}_{\max_A} = \frac{\sup_S d(\boldsymbol{x}|\xi^*_{A_2})}{\sup_{\Gamma_0} \sup_S d(\boldsymbol{x}|\Upsilon, \xi^*_{A_1})} \quad \text{and} \quad \text{eff}_{\text{avg}_A} = \frac{\int_S d(\boldsymbol{x}|\xi^*_{A_2})d\boldsymbol{x}}{\sup_{\Gamma_0} \int_S d(\boldsymbol{x}|\Upsilon, \xi^*_{A_1})d\boldsymbol{x}},$$

where $d(\boldsymbol{x}|\xi^*_{A_2}) = g(\boldsymbol{x})' \widetilde{M}_1^{-1}(\xi^*_{A_2})g(\boldsymbol{x})$. Other efficiencies $\text{eff}_{\max_D}$, $\text{eff}_{\text{avg}_D}$, $\text{eff}_{\max_G}$, $\text{eff}_{\text{avg}_G}$, $\text{eff}_{\max_Q}$ and $\text{eff}_{\text{avg}_Q}$ are similarly defined. In Figures 1 and 2 the efficiencies are plotted against ν for $k=2$. Plots for $k \geq 3$ reveal similar trends except that the range of ν for which each efficiency is higher than 1 is wider. G- and Q- optimal MMSE invariant first order designs perform as good as the corresponding optimal second order designs for a wide range of ν. However, $\text{eff}_{\max_D}$ and $\text{eff}_{\text{avg}_D}$ decrease rapidly. This is mainly because ν is not reflected in D-optimal MMSE invarinat first order designs. Consequently efficiency of D-optimal MMSE invariant first order designs to D-optimal second order designs is very sensitive to ν. When a decision is made on whether we will use a MMSE first order design or an optimal second order design, relatively more accurate estimate of ν is required for D-optimality than for other optimalities.

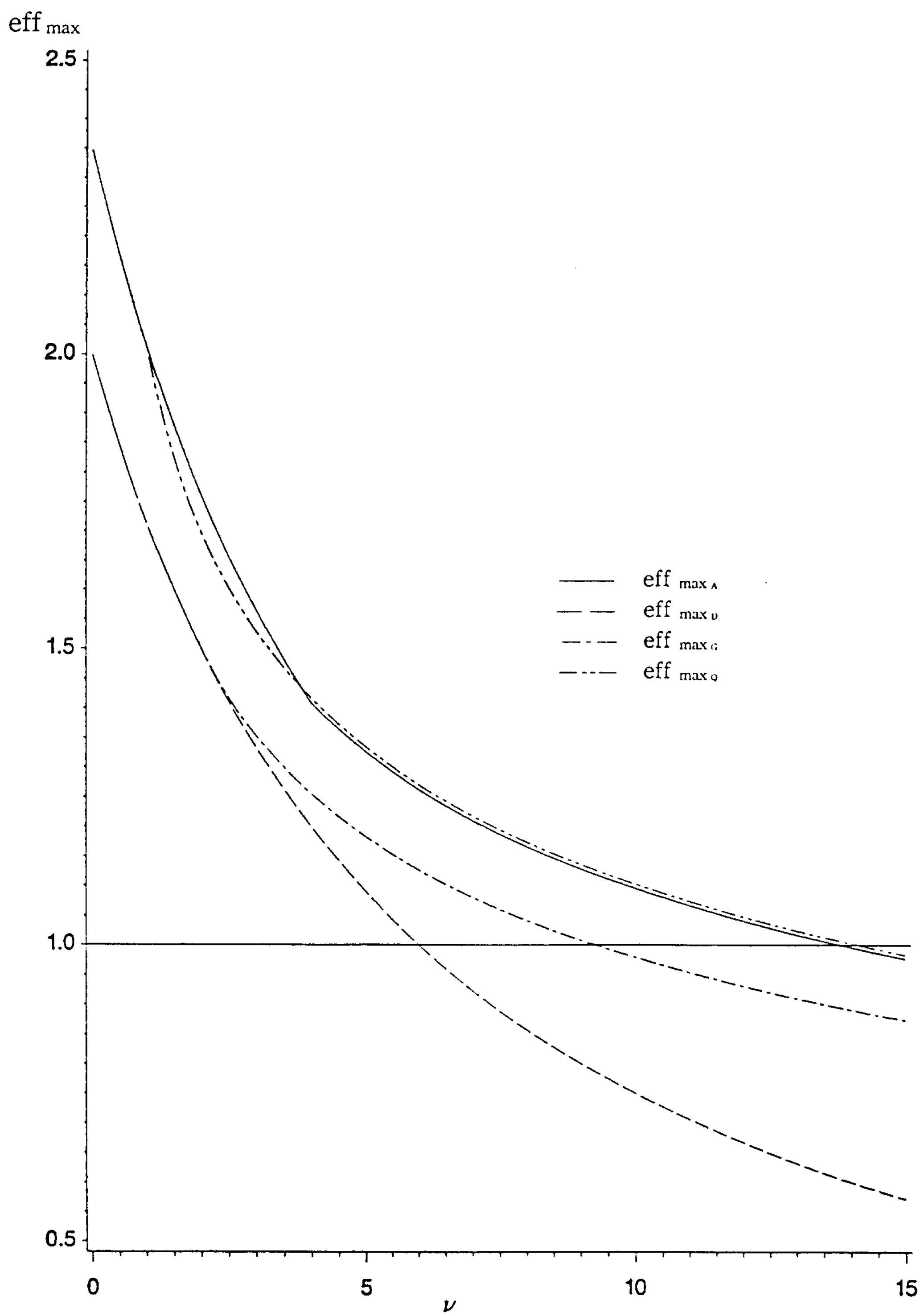

Figure 1. Plot of eff_{max} for $k = 2$.

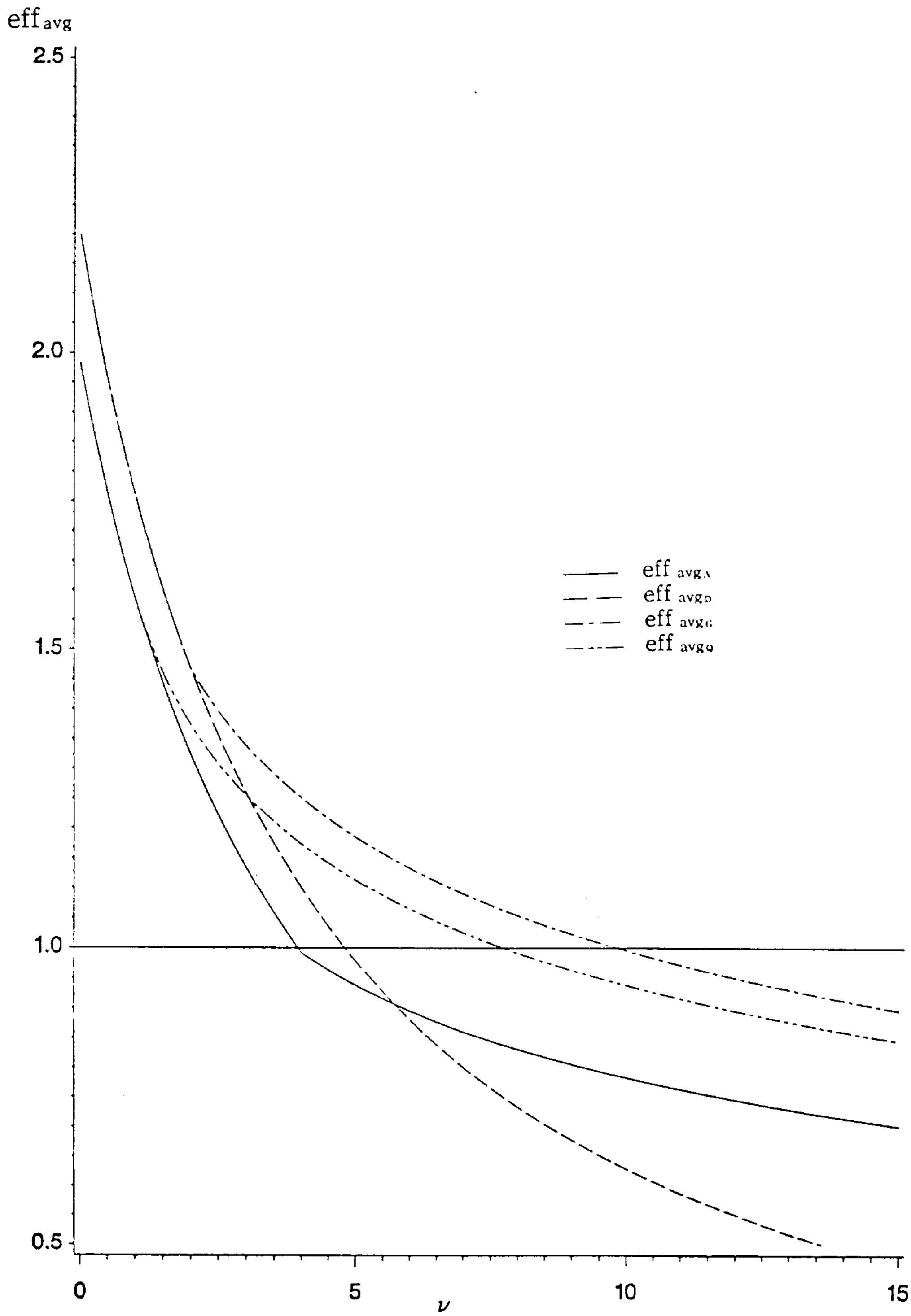

Figure 2. Plot of eff_{avg} for $k = 2$.

References

[1] Box, G.E.P. and Draper, N. (1959). A Basis for the Selection of A Response Surface Design, *Journal of American Statistical Association*, Vol. 54, 622-654.

[2] Box, G.E.P. and Draper, N.R. (1963). The Choice of A Second Order Rotatable Design, *Biometrika*, Vol. 50, 335-352.

[3] Draper, N.R. and Lawrence, W.E. (1965). Designs Which Minimize Model Inadequacies: Cuboidal Regions of Interest, *Biometrika*, Vol. 52, 111-118.

[4] Draper, N.R. and Sanders, E.R. (1988). Designs for Minimum Bias Estimation, *Technometrics*, Vol. 30, 319-325.

[5] Heiligers, B. (1991). Admissibility of Experimental Designs in Linear Regression with Constant Term, *Journal of Statistical Planning and Inference*, Vol. 28, 107-123.

[6] Heiligers, B. (1992). Admissible Experimental Designs in Multiple Polynomial Regression, *Journal of Statistical Planning and Inference*, Vol. 31, 219-233.

[7] Heiligers, B. and Schneider, K. (1992). Invariant Admissible and Optimal Designs in Cubic Regression on the ν-ball, *Journal of Statistical Planning and Inference*, Vol. 31, 113-125.

[8] Huber, J. (1975). Robustness and Designs, In: Srivastava, J.N. ed., *A Survey of Statistical Design and Linear Models,* North Holland, Amsterdam, 287-303.

[9] Karson, M.J., Manson, A.R. and Hader, R.J. (1969). Minimum Bias Estimation and Experimental Design for Response Surfaces, *Technometrics*, Vol. 11, 461-475.

[10] Kiefer, J. (1959). Optimal Experimental Desings (with discussion), *Journal of Royal Statistical Society Ser. B*, Vol. 21, 272-319.

[11] Li, K.C. (1984). Robust Regression Designs When the Design Space Consists of Finitely Many Points, *Annals of Statistics*, Vol. 12, 269-282.

[12] Li, K.C. and Notz, W. (1982). Robust Designs for Nearly Linear Regression, *Journal of Statistical Planning and Inference*, Vol. 6, 135-151.

[13] Markus, M.B. and Sacks, J. (1976). Robust Designs for Regression Problems, In: Gupta, S.S. and Moore, D.S. eds., *Statistical Theory and Related Topics II*, Academic Press, New York, 245-268.

[14] Myers, R.H. and Lahoda, S.J. (1975). A Generalization of the Response Surface Mean Squared Error Criterion with A Specific Application to the Slope, *Technometrics*, Vol. 17, 481-486.

[15] Park, J.Y. (1990). Designs for Estimating the Difference Between Two Responses, *Communications in Statistics-Theory and Methods*, Vol. 19, 4773-4787.

[16] Pesotchinsky, L. (1982). Optimal Robust Designs: Linear Regression in R^k, *Annals of Statistics*, Vol. 10, 511-525.

[17] Sacks, J. and Ylvisaker, D. (1978). Linear Estimation for Approximately Linear Models, *Annals of Statistics*, Vol. 6, 1122-1137.

[18] Vining, G.G. and Myers, R.H. (1991). A Graphic Approach for Evaluating Response Surface Designs in Terms of the Mean Squared Error of Prediction, *Technometrics*, Vol. 33, 315-326.

[19] Weins, D.P. (1992). Minimax Designs for Approximately Linear Regression, *Journal of Statistical Planning and Inference*, Vol. 31. 353-371.